3D FACIAL VISUALIZATION THROUGH ADAPTIVE SPREAD SPECTRUM SYNCHRONOUS SCALABLE (A4S) DATA HIDING

K. Hayat[a], W. Puech[a], G. Gesquiere[b] and G. Subsol[a]

[a] LIRMM, UMR CNRS 5506, University of Montpellier II, FRANCE
[b] LSIS, UMR CNRS 6168, Aix- Marseille University, FRANCE
khizar.hayat@lirmm.fr, william.puech@lirmm.fr, gilles.gesquiere@lsis.org, gerard.subsol@lirmm.fr

ABSTRACT

An adaptive spread spectrum synchronous scalable(A4S) data hiding strategy is being put forward to integrate the disparate 3D facial visualization data, into a single JPEG2000 format file with the aim to cater diverse clients in various bandwidth scenarios. The method is both robust and imperceptible in the sense that the robustness of the spread spectrum (SS) is coupled with the removable embedding that ensures highest possible visualization quality. The SS embedding of the DWT-domain 2.5D facial model is carried out in the transform domain YCrCb components, of the 2D texture, from the coding stream of JPEG2000 codec just after the DWT stage. High depth map quality is ensured through the adaptation of synchronization during embedding that would exclude some highest frequency subbands. The results show that the method can be exploited for video-surveillance and video-conference applications.

1. INTRODUCTION

Transmitting digital 3D face data in real-time has been a research issue for quite a long time. When it comes to the real-time, two main areas, viz. conferencing and surveillance, suddenly come to mind. In the earlier video-conference applications, the aim was to change the viewpoint of the speaker. Despite the fact that many technological barriers have been eliminated, thanks to the availability of cheap cameras, powerful graphic cards and high bitrate networks, there is still no efficient product that offers a true conferencing environment. In fact, the transmission of only a small number of parameters of movement or expression can materialize the video through low speed networks. In addition recent technology innovations have increased the bandwidth of conventional telephone lines to several Mbps. All this notwithstanding, the bitrate limitation still exists in the case of many devices like PDA or mobile phones. The matter becomes even more critical, in particular, in remote video-surveillance applications which are gaining increasing economic importance. Some companies offer to send surveillance images on the mobile phones/PDAs of authorized persons but these are only 2D images whereby the identification of persons is very difficult, especially in poor light conditions.

The objective of our work is to reduce the data considerably for optimal real-time 3D facial visualization in a client/server environment. A typical 3D visualization is based on a 2D intensity image, called texture and a corresponding 3D shape rendered in the form of a range image. A range image, also sometimes called a depth image, is an image in which the pixel value reflects the distance from the sensor to the imaged surface [1]. In essence, for 3D visualization, one would thus have to manipulate at least two files. We propose to unify these two data into a single standard JPEG2000 format file. The apparent advantage of this strategy is that the multiresolution nature of wavelets would offer the required scalability which is necessary for the client diversity. Wavelets have been extensively employed [2] for face related applications but rather than the visualization, the focus had traditionally been on feature extraction for face recognition. Moreover, to cater for the diverse clients and unify the disparate 3D visualization data we are proposing in this work an adaptive spread spectrum synchronous and scalable (A4S) data hiding strategy which is simultaneously scalable. For the unification of the 2D texture and 2.5D model, the A4S data hiding strategy is being put forward wherein the 2.5D data is embedded in the corresponding 2D texture in the wavelet transform domain.

2. WAVELET-BASED DATA HIDING

Data hiding methods for JPEG2000 images must process the code blocks independently [3]. That is why a majority of the classical wavelet-based data hiding methods proposed in the literature have not been compatible with the JPEG2000 scheme. The JPEG2000-based image authentication method of [4] employs extended scalar quantization and hashing for the protection of all the coefficients of the wavelet decomposition. The process involves feature extraction by wavelets to result in digital signature which, after encryption and error correction coding, is embedded as a removable watermark using the well-known quantization index modulation technique called dither modulation. One blind method [5] transforms the original image by one-level wavelet transform and sets the three higher subbands to zero before inverse transforming it to get the modified image. The difference values between

the original image and the modified image are used to ascertain the potential embedding locations of which a subset is selected pseudo-randomly for embedding.

The SS method in [6] embeds watermark information in the coefficients of LL and HH subbands of different decompositions. Traditionally, correlation analysis has been an integral part of the SS methods reported in various works. Based on the significant difference of wavelet coefficient quantization, a blind algorithm [7] groups every seven non-overlap wavelet coefficients of the host image into a block. Uccheddu *et al.* [8] adopt a wavelet framework in their blind watermarking scheme for 3D models under the assumption that the host meshes are semi-regular, thus paving the way for a wavelet decomposition and embedding of the watermark at a suitable resolution level. Agreste *et al.* [9] put forward a strong wavelet-based watermarking algorithm which embeds the watermark into high frequency DWT components of a specific sub-image and it is calculated in correlation with the image features and statistical properties. Watermark detection applies a re-synchronization between the original and watermarked image involving the Neyman-Pearson statistic criterion for correlation.

3. THE PROPOSED A4S DATA HIDING METHOD

In this section, we present our method for an adaptive scalable transfer and online visualization of 3D faces. For the embedding, it is proposed over here to employ a spread spectrum (SS) data hiding strategy. The SS methods offer high robustness at the expense of cover quality but this quality loss is reversible, in our case, since the embedded data can be removed after recovery. The proposed adaptive synchronization is helpful in improving the quality of the range data approximation for a given texture approximation. Suppose a $N \times N$ texture image has a depth map of $m \times m$ coefficients. In the spatial domain, let each of the coefficient corresponds to a $t \times t$ pixel block of the related texture, where $t = \frac{N}{m}$. Suppose the texture is to be JPEG2000 coded at DWT decomposition level L, implying $R = L + 1$ resolutions. Let us apply lossless DWT to the range coefficients at level L', where $L' \leq L$. For embedding we interrupt the JPEG2000

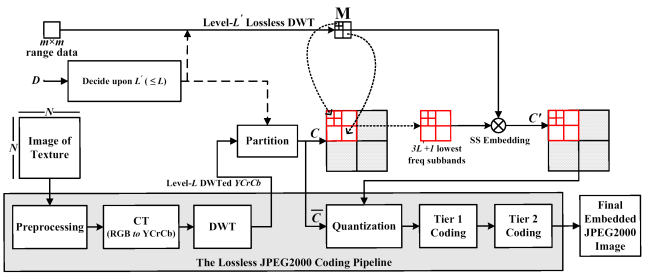


Fig. 1: Global overview of the encoding process.

coding, of the texture, after the DWT step, as illustrated in Fig. 1, and get one of the transformed YCrCb components as a carrier for embedding. The carrier (C) is partitioned into $m \times m$ equal-sized blocks, $B_{i,j}$, with size dependent on the value of L'. If $L' = L$ then C consists of whole of the selected component(s) and embedding block size remains $t \times t$, since no subband is excluded from the possible data insertion. Otherwise, for $L' < L$, only a subset subbands - the lowest $3(L - L') + 1$ of the original $3L + 1$ after excluding the remaining $3L'$ higher frequency subbands - constitute C and $B_{i,j}$ has a reduced size of $t/2^{(L-L')} \times t/2^{(L-L')}$. Care must be taken of the fact that block size must be large enough to reliably recover the embedded data after correlation. Hence the important factors in reaching a decision is based on the value of L', the block size and the involvement or otherwise of more than one YCrCb component in embedding.

The process of embedding is done in the DWT-domain of the carrier, C, which is one or more of the YCrCb components of the transformed texture. The criteria for ascertaining the carrier depends on L', t and ω, where ω is the number of bits assigned to represent a single DWT domain range data coefficient.

Algorithm 1 Embedding of the range data coefficient $d_{i,j}$ in the corresponding block partition $B_{i,j}$.

begin
get the (i, j)th partition $B_{i,j}$ of the cover and the corresponding s-bit coefficient $d_{i,j}$
partition $B_{i,j}$ to s sub-blocks, $b_0, b_1, \ldots b_{s-1}$
for $k \leftarrow 0$ *to* $s - 1$ **do**
  read the kth bit β_k of the DEM coefficient $d_{i,j}$
  if $\beta_k = 0$ **then**
    set $b'_k \leftarrow b_k - W$
  else
    set $b'_k \leftarrow b_k + W$
  end if
  replace b_k by b'_k in the block $B_{i,j}$
end for
replace $B_{i,j}$ by $B'_{i,j}$
end

The choice whether to embed in Y or Cr/Cb plane depends on the fact that Y plane embedding would distort the encoded image while the chrominance plane embedding would escalate the final file size. Neither is the former a serious issue, as embedding is removable, nor is the latter, since it may matter only when $L' = L$. The embedding process of a DEM coefficient in a given block (size $= t^2/(2^{(L-L')})^2$) is elaborated by the logic outlined in algorithm 1, where W is a one-time pseudo-randomly constructed matrix. Since one coefficient is embedded per block, each $B_{i,j}$ is repartitioned into as many sub-blocks (b_k) as the number of bits used to represent a single transformed coefficient. Each of b_k from $B_{i,j}$ would carry the k^{th} bit β_k of the transformed range depth coefficient $d_{i,j}$. Embedding depends on a key generating a pseudo-random matrix W with entries from the set $\{\alpha, -\alpha\}$. The matrix

W has the same size as of b_k, i.e. $W = [y_{uv}]_{n\times n}$, where $y_{uv} \in \{\alpha, -\alpha\}$. The scalar, α, is referred to the strength of embedding. If the bit to embed, β_k, is '1', then matrix W is added to the matrix b_k, otherwise W is subtracted from b_k. The result is a new matrix b'_k which replaces b_k as a sub-block in the embedded block, $B'_{i,j}$, of the marked carrier C'. At the decoding end, the quality of the range data would depend on the difference between L and L'. The larger the difference, $(L - L')$, higher will be the quality and vice versa.

Algorithm 2 Algorithm for the recovery of the range data coefficient $d_{i,j}$ and the corresponding block $B_{i,j}$

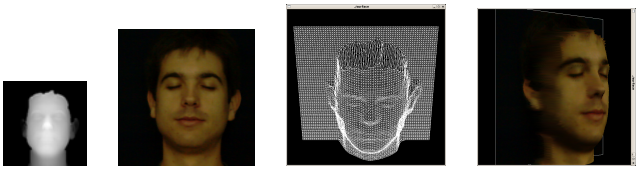
begin
get the (i, j)th partition $B'_{i,j}$ of the cover
partition $B'_{i,j}$ to s sub-blocks, $b'_0, b'_1, \ldots b'_{s-1}$
set $d_{i,j} = 0$
for $k \leftarrow 0$ *to* $s - 1$ **do**
  Compute the Pearson's correlation coefficient ρ between the vectors b'_k and W
  if $\rho < \tau$, where τ is a threshold **then**
    set$\beta_k \leftarrow 0$
    set $b_k \leftarrow b'_k + W$
  else
    set $\beta_k \leftarrow 1$
    set $b_k \leftarrow b'_k - W$
  end if
  replace kth bit of $d_{i,j}$ by β_k
  replace b'_k by b_k in the block $B'_{i,j}$
end for
replace $B'_{i,j}$ by $B_{i,j}$
end

The above coded image can be utilized like any other JPEG2000 image and sent across any communication channel. The decoding is more or less converse to the above process. Just before the inverse DWT stage of the JPEG2000 decoder the range data can be extracted from C', using the above mentioned partitioning scheme, i.e. $B'_{i,j}$ blocks and their b'_k sub-blocks. Algorithm 2 shows the flowchart for the recovery of a range depth coefficient $d_{i,j}$ from sub-blocks b'_k of $B'_{i,j}$ and eventual reconstruction of $b_{i,j}$ and ultimately $B_{i,j}$. A given k^{th} sub-block b'_k and also the matrix W can be treated as a column/row vector and the Pearson's correlation coefficient, ρ, can be computed thereof. If ρ is less than certain threshold, τ then the embedded bit β was a '0', otherwise it was a '1'. Once this is determined then it is obvious that what were the entries of b_k, i.e. if β is '0' then add W to b'_k, otherwise subtract W from b'_k.

Next comes the reconstruction phase wherein by the application of 0-*padding* one can have $L+1$ and $L'+1$ different approximation images of texture and its range data, respectively. And this is where one achieve the scalability goal. Our method is based on the assumption that it is not necessary that all the subbands are available for reconstruction, i.e. only a part of C' is on hand. This is one of the main advantages of the method since the range data and the texture can be reconstructed with even a small subset of the coefficients of the carrier. The resolution scalability of wavelets and the synchronized character of our method enable a $3D$ visualization even with fewer than original resolution layers as a result of partial or delayed data transfer. The idea is to have a $3D$ visualization utilizing lower frequency subbands at level L'', say, where $L'' \leq L$. For the rest of $L - L''$ parts one can always pad a 0 for each of their coefficient. The inverse DWT of the 0-stuffed transform components will yield what is known as image of approximation of level L''. Before this, data related to the third dimension must be extracted whose size depends on both L'' and L'. Thus if $L' \leq L''$ one will always have the entire set of the embedded depth coefficient since all of them will be extractable. We would have a level 0 approximate final range map after inverse DWT, of the highest possible quality. On the other hand if $L' > L''$, one would have to pad 0's for all coefficients of higher $L' - L''$ subbands of transformed range map before inverse DWT that would result in a level L''-approximate range data of an inferior quality.

4. APPLICATION TO 3D FACE VISUALIZATION

We applied our method to more than 10 examples from FRAV3D[1] database, each consisting of a 512×512 face texture and a corresponding 64×64 depth map of 8-bit coefficients - one of the examples is given in Fig. 2. The proportion between the size of the face texture and the depth map suggests that one 2.5D coefficient corresponds to a 8×8 texture block. Hence, even if we involve all the three planes (i.e. YCrCb) in embedding, we can adapt by one level, at most, with a SS embedding. The 8 bit depth map was subjected to level-L' lossless DWT before embedding in the three DWT-domain components from the JPEG2000 coder. Taking $L = 4$ would imply $L' = 4$ for synchronous embedding and $L' = 3$ for a one level adaptation in synchronous embedding. To embed one bit, the former would manipulate about 21 while the latter about 5 coefficients of Y, Cr or Cb plane in the DWT domain.



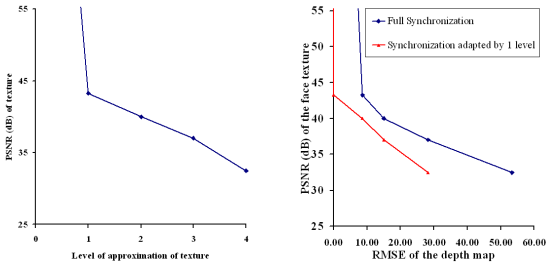
(a) (b) (c) (d)

Fig. 2: Example face visualization: a) 64×64 depth map, b) 512×512 pixel face texture, c) The corresponding 3D mesh, d) A 3D view.

For the set of face examples, the maximum of the lowest value of α needed for the 100% recovery of the embedded

[1] http://www.frav.es/databases/FRAV3d/

data was found to be 26 for the synchronous case and 45 for the adaptive synchronous case. The resultant watermarked JPEG2000 format image was degraded to a mean PSNR of 17.15 dB for the former and a mean PSNR of 15.16 dB for the latter. Even with that much distortion, the reversible nature of the method enabled us to fully recover the embedded data on one hand and achieve the maximum possible quality in case of the texture, i.e. with a PSNR of infinity. Fig. 3.a plots the mean PSNR against the level of approximation for the reconstruction of texture. It can be seen that even a texture reconstructed with as low as 0.39% of the transmitted coefficients has a mean PSNR of around 30 dB. The advantage of adaptation is evident from the fact that, as shown in Fig. 3.b, for the same texture quality, even one level adaptation offers far better depth map quality as compared to full synchronization.

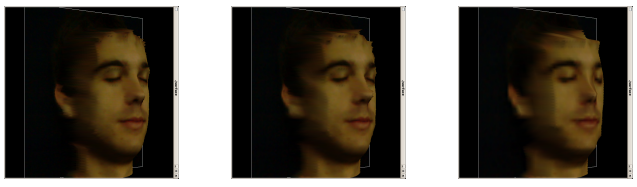


(a) (b)

Fig. 3: Variation in the texture quality as a function of: a) its approximation level, b) RMSE of the extracted depth map.

(a) Level-1 (b) Level-2 (c) Level-3

Fig. 4: 3D visualization corresponding to Fig. 2 realized through the overlaying of **level-**l approximate texture on the extracted **level-**l approximate depth map.



(a) Level-1 (b) Level-2 (c) Level-3

Fig. 5: 3D visualization corresponding to Fig. 2 realized through the overlaying of **level-**l approximate texture on the extracted **level-**$(l-1)$ approximate depth map.

The evident edge of the adaptive strategy in depth map quality is conspicuous by the comparison of 3D visualizations illustrated in Fig. 4 and Fig. 5. The latter is always better by one level (four times, roughly) as far as depth map quality is concerned. Hence with the same number of transmitted coefficients, one gets the same level-l approximate texture for the two cases but the one level adaptive case would have level-$l-1$ approximate depth map as compared to the relatively inferior level-l approximate depth map for the fully synchronized case.

5. CONCLUSION

The proposed A4S data hiding approach has the peculiarity of being robust and imperceptible, simultaneously. With the adaptation in synchronization, higher quality of the depth map was ensured but the extent of adaptation is strictly dependent on the embedding factor. The quality offered by the method, for various approximations, is the ultimate as 100% of the texture and depth map coefficients are recoverable. The obtained results have been interesting in the sense that even with a small difference in the sizes of the depth map and the 2D face image one got a satisfactory 3D visualization. The trend of our results implies that an effective visualization is possible from even a texture reconstructed with as low as 0.39% of the transmitted coefficients having a PSNR of around 30 dB. The advantage of adaptation is evident from the fact that for the same texture quality, even one level adaptation improved the 2.5D quality a lot.

6. REFERENCES

[1] K.W. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches and Challenges in 3D and Multi-modal 3D + 2D Face Recognition," *Computer Vision & Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[2] D.-Q. Dai and H. Yan, *Face Recognition*, chapter Wavelets and Face Recognition, I-Tech Education and Publishing, Vienna, Austria, 2007.

[3] P. Meerwald and A. Uhl, "A Survey of Wavelet-Domain Watermarking Algorithms," in *Proc. SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents III*, San Jose, CA, USA, January 2001, vol. 4314, pp. 505–516, SPIE, IS&T.

[4] M. Schlauweg, D. Pröfrock, and E. Müller, "JPEG2000-Based Secure Image Authentication," in *MM&Sec '06: Proceedings of the 8th Workshop on Multimedia and Security*, New York, NY, USA, 2006, pp. 62–67, ACM.

[5] J. L. Liu, D. C. Lou, M. C. Chang, and H. K. Tso, "A Robust Watermarking Scheme Using Self-Reference Image," *Computer Standards & Interfaces*, vol. 28, pp. 356–367, 2006.

[6] Santi P. Maitya, Malay K. Kundub, and Tirtha S. Das, "Robust SS Watermarking with Improved Capacity," *Pattern Recognition Letters*, vol. 28, no. 3, pp. 350–356, 2007.

[7] W.-H. Lin, S.-J. Horng, T.-W. Kao, P. Fan, C.-L. Lee, and Y. Pan, "An Efficient Watermarking Method Based on Significant Difference of Wavelet Coefficient Quantization," *IEEE Trans. Multimedia*, vol. 10, no. 5, pp. 746–757, 2008.

[8] F. Uccheddu, M. Corsini, and M. Barni, "Wavelet-Based Blind Watermarking of 3D Models," in *MM&Sec '04: Proceedings of the 2004 workshop on Multimedia and security*, New York, NY, USA, 2004, pp. 143–154, ACM.

[9] S. Agreste, G. Andaloro, D. Prestipino, and L. Puccio, "An Image Adaptive, Wavelet-Based Watermarking of Digital Images," *Journal of Computational and Applied Mathematics*, vol. 210, no. 1–2, pp. 13–21, 2007.